Tutor/s

Dr. Oscar Núñez Burcio
Departament d' Enginyeria Química i Química Analítica

Dr. Javier Saurina Purroy
Departament d' Enginyeria Química i Química Analítica

Grau de Química

Treball Final de Grau

HPLC-FLD and HPLC-UV Fingerprinting for the Characterization, Classification and Authentication of Paprika.

Caracterización, Clasificación y Autenticación de Pimentones mediante huellas HPLC-FLD y HPLC-UV.

Luis R. Rodríguez Javier

January 2021

Universitat de Barcelona

BKC Barcelona Knowledge Campus

No human investigation can be called real knowledge if it does not pass through mathematical demonstrations.

Leonardo da Vinci

I would like to thank my supervisors in this work, Oscar Núñez and Javier Saurina, that have guided and helped me during the whole project while giving me the space to grow as a chemist, allowing me to improve my liquid chromatography and chemometrics skills and knowledge.

I would also like to thank all the members of CECEM for accepting me with open arms, especially to Guillem, Nerea, and Josep, for being there when I needed them the most.

I can't forget my friends, family, and college classmates, who have made me who I am today. In particular, I would like to thank Aïda Valverde and Victor Rubio, who have accompanied me during these months, motivating me to move on, not only in the project but also in the day-to-day.

REPORT

CONTENTS

1. SUMMARY

In recent years, there has been a growing interest in the quality of natural foods from society. Consequently, the demand for food traceability has increased and, along with it, measures to satisfy consumer concerns. One of these tools is the Protected Designation of Origin (PDO), which guarantees the quality and geographical origin of a product. In the case of paprika, a red spice obtained after drying and grinding certain varieties of red peppers, there are seven PDOs in the European Union. Even with this quality seal, the products do not escape from food fraud like adulterations with chemicals or paprika from other geographical origins. To avoid this, it is essential to develop new techniques to solve these problems.

This project aims to create a non-targeted method that allows the classification of paprika according to their production region. To this end, chromatographic fingerprints obtained by high-performance liquid chromatography with fluorescence (HPLC-FLD) and ultraviolet (HPLC-UV) detection were proposed as chemical descriptors. The chromatographic separation was performed on a C18 reversed-phase column under gradient elution using acidified water (0.1% formic acid) and acetonitrile as the mobile phase components.

The obtained chromatographic fingerprints were analyzed using chemometric techniques such as principal component analysis (PCA) and partial least squares - discriminant analysis (PLS-DA). In these studies, apart from ensuring that the proposed methods were robust and reproducible, a classification of the samples with 95.8% accuracy was obtained for both HPLC-UV and HPLC-FLD fingerprints.

Keywords: Paprika, Food fraud, PDO, fingerprints, HPLC-FLD, HPLC-UV, PCA, PLS-DA.

2. RESUMEN

En los últimos años, ha habido un interés creciente en la calidad de los alimentos naturales por parte de la sociedad. En consecuencia, ha aumentado la demanda de la trazabilidad alimentaria y, junto a ella, medidas para saciar la preocupación de los consumidores. Una de estas herramientas es la designación de Denominación de Origen Protegido (DOP), la cual permite garantizar la calidad y el origen geográfico de un producto. En el caso del pimentón, una especia roja obtenida tras secar y triturar ciertas variedades de pimientos, se encuentran siete DOP en la Unión Europea. Aún con este sello de calidad, los productos no se libran del fraude alimentario, llegando a encontrar adulteraciones con químicos o alimentos de otras regiones. Para evitarlo, es importante el desarrollo de nuevas técnicas que se adapten a cada uno de los problemas.

Este proyecto tiene como objetivo la creación de un método no dirigido que permita la clasificación de pimentones en función de su región de producción. Con este fin, se han propuesto como descriptores químicos las huellas cromatográficas obtenidas mediante cromatografía líquida de alta eficacia con detección de fluorescencia (HPLC-FLD) y ultravioleta (HPLC-UV). La separación cromatográfica se ha llevado a cabo con una columna de fase invertida C18 con gradiente de elución utilizando agua acidificada (0.1% de ácido fórmico) y acetonitrilo como fase móvil.

Las huellas cromatográficas obtenidas han sido analizadas empleando técnicas quimiométricas como el análisis de componentes principales (PCA) y el análisis discriminante por regresión de mínimos cuadrados parciales (PLS-DA). En estos estudios, aparte de confirmar que los métodos propuestos son robustos y reproducibles, se ha obtenido una clasificación de las muestras con un 95.8% de éxito para ambas huellas HPLC-UV y HPLC-FLD.

Palabras clave: Pimentón, Fraude alimentario, DOP, Huella cromatográfica, HPLC-FLD, HPLC-UV, PCA, PLS-DA.

3. INTRODUCTION

3.1. FOOD AUTHENTICATION

In the last years, the consumption of natural products instead of processed ones has increased significantly. Even though there has never been so much interest in healthy food as there is today, consumers still increase the demand about information, either to avoid certain allergens, to have a balanced diet, or to know the origin [1].

The public awareness about food safety and quality causes the constant development in food authentication, which is the process of verifying that the product is in compliance with the label description, including among others, the origin (geographical, species or genetic), production method, and processing technologies [2].

One of the biggest problems in this situation is food fraud, defined as a term that encompasses the deliberate addition, substitution, misrepresentation or tampering of a product for economic benefit at the expense of the consumer [3]. Natural products are prone to be adulterated with more economic additives, and these adulterants range from natural products with similar properties to harmful chemicals. Some examples are the adulteration of extra virgin olive oil with refined oils [4], Sudan dyes in spices [5] or the melamine adulteration of milk powder [6]. Although they are not equated, these previous situations are considered a public health risk and, for that reason, it is important that the food industry has a good traceability, quality control, and security to ensure that the products satisfy consumer expectations.

In order to control these measures over the world, one of the many options are the regulations made by the International Standardization Organization (ISO) such as ISO:9001 (quality management) or ISO:22005:2007 (traceability in the food chain). Furthermore, the European Union has its own institution for food legislation. The European Food Safety Authority (EFSA) includes a set of laws and guidelines to assure food quality and safety [7]. With the aim of controlling the product precedence as well as to support the good practices in rural and agricultural activities, the Council Regulation (EC) No 510/2006 [8] proposes, among others, the Protected Designation of Origin (PDO), which relates the products to the geographical area

where they are prepared, processed and produced. Even so, these products are not spared from food fraud and authentication has an important role preventing it.

3.2. PAPRIKA

Paprika, or chili pepper, is a red powder spice obtained after drying and grinding some varieties of red peppers of the genus *Capsicum*, which belongs to the Solanaceae family [9]. Due to its characteristic color and taste, paprika is commonly used to add flavor and color to various ethnic dishes and many foods such as soup, beverages, drinks, meat, baked goods, candy, ice cream, and seasoning mixes [10], although others applications as cosmetics, personal protection sprays, adsorbents to remove contaminants, and medicine are reported [11,12].

The pepper used for the preparation of paprika has its origin in the Mexican Indians' diet, with approximately 39 wild, semidomestic, and domestic species within the genus *Capsicum*, the most common is the *Capsicum annuum* [11,13]. With the marine exploration period at the end of the XV century, these seeds arrived to Europe and their farming was adapted to the conditions of each region. Since then, the globalization of paprika increased exponentially due the ease of hybridization [11]. Nowadays, *C. annuum* is one of the most cultivated vegetables in the world [14].

In Europe seven Protected Designation of Origin for paprika exist: Slovakia Žitavská paprika (Slovakia), Kalocsai fűszerpaprika-őrlemény and Szegedi fűszerpaprika-őrlemény (Hungary), Piment d'Espelette (France), Pimentón de la Vera (Extremadura, Spain), Pimentón de Murcia (Murcia, Spain), and since 2020, Pebre bord de Mallorca (Mallorca, Spain) [15]. It is obvious to think that the paprika from two regions would taste different due to natural and human factors such as the climatic condition, cultivation area and techniques, water resources, soil management or manufacturing practices, but properties and organic compounds of paprika are also specific for each region [16].

Regarding the flavor variety, three important groups can be distinguished: hot paprika, sweet paprika, and bittersweet paprika. Besides, during the elaboration of the paprika, in some cases peppers are dried by employing oak wood smoke, adding three new important flavors: smoked hot paprika, smoked sweet paprika, and smoked bittersweet paprika [17].

Concerning the organic compounds present in paprika, flavonoids and carotenoids are responsible for the characteristic red color [11] while capsaicinoids and capsinoids grant the pungency (spicy flavor) and some health benefits along with other bioactive compounds [18]. Some examples are ascorbic acid (vitamin C) or tocopherols (vitamin E), but among the most important ones are phenolic and polyphenolic compounds, due to their antioxidant activity that can promote vascular protection, antiosteoporotic, anti-inflammatory, antitumor, antiobesity, antiallergic and analgesic effects amidst others [12,16,19].

Apart from its great utility in health, with the advance of the analytical methodologies, polyphenol content and distribution have been used for the evaluation of food quality and the detection of frauds in paprika [18].

3.3. ANALYTICAL METHODOLOGIES FOR FOOD FRAUD DETECTION

With the increasing food fraud complexity, an improvement in the analytical approaches is required. In food authentication there are several methodologies based on infrared, Raman, and nuclear magnetic resonance (NMR) spectroscopies, isotopic analysis by nuclear magnetic resonance, chromatographic techniques, fluorescent spectroscopy, ultraviolet-visible (UV-VIS) spectroscopy, and DNA-based technology among others [20]. Nowadays, liquid chromatography with ultraviolet detection (LC-UV) or coupled to mass spectrometry (LC-MS) or high-resolution mass spectrometry (LC-HRMS) are some of the most useful techniques to detect fraud [21].

Regarding paprika, studies with UV-VIS [22], visible - near infrared [23], Fourier transform mid-infrared [24], Raman [25], NMR [26], and energy dispersive X-ray fluorescence [15] spectroscopies, as well as liquid chromatography with electrochemical detection using screen-printed carbon-based nanomaterial electrodes [9], LC-UV [27], LC-MS [12], and LC-HRMS [13] have been employed to assess their characterization, classification and authentication.

As seen in these works, regardless the instrumentation, two analytical strategies can be mainly employed: targeted and non-targeted methodologies. The percentage of food authentication works using those approaches has increased by 300% between 2007 and 2016 due the rising demand in our society to guarantee food authenticity [28].

3.3.1. Targeted analysis

Targeted analysis is an analytical strategy whose objective is the detection or quantification of pre-defined target compounds. When these targets are linked to the authentication issue, they are called analytical markers and two types can be distinguished. Primary markers refer to those whose result is directly related to the authentication, often based on specific legal limits. Secondary markers are those indirectly related, for example, those associated with the geographic origin, specie or agricultural production methods [28]. In the case of paprika these markers usually are polyphenols or capsaicinoids [13].

Targeted methods are often quantitative and have a greater sensitivity and selectivity respect to non-targeted methods, but, as only one or a few analytes are studied, the information offered to protect the consumer against adulteration may be limited. This strategy is usually used when the suspected target is a primary marker, otherwise non-targeted analysis is preferably employed.

3.3.2. Non-targeted analysis

Non-targeted analysis, also known as untargeted analysis or fingerprinting, relies on complex instrumental data that can be generated with a wide variety of analytical methodologies whose objective is to obtain as many features (peaks or signals) of metabolites as possible without the requirement of identifying the compounds [29]. These methods usually are qualitative and no markers are defined.

Food discrimination and classification are carried out contrasting and comparing patterns between the obtained fingerprints. This pattern represents the metabolites that change in response to natural or artificial alterations, for example, geographical origin or adulteration [29]. In order to avoid a loss of signals or information, it is important to use sample treatments as simple and unspecific as possible [30]. The ability to detect multiple small variations in the food product makes untargeted methods a perfect option to deal with the complexity of modern food fraud, especially when no primary or secondary markers are available [28].

The principal disadvantage of these methodologies arises in the large amount of chemical data obtained. First, a data matrix is built converting the fingerprints into valuable information with a specific software. Then, all this data is analyzed with a multivariate statistical program, so, in order to prevent food fraud, a chemometric study has to be done.

3.4. CHEMOMETRICS

Chemometrics is defined by the International Chemometrics Society as "the chemical discipline that uses mathematical, statistical, and other methods employing formal logic to design or select optimal measurement procedures and experiments, and to provide maximum relevant chemical information by analyzing large amounts of chemical data". Many applications of chemometrics have been proposed in the last decade to solve the adulteration and food fraud problematic due the advantages of multivariate statistical methods [31].

One of the main classifications in chemometric distinguishes between supervised and non-supervised methods. Non-supervised methods consist of identifying groups or connections between samples without any prior knowledge of the class or group to which they belong. Instead, supervised methods require previous information, such as the belonging class, and uses it as a reference to predict the results. Two representative techniques of these methods are principal component analysis (PCA) and partial least squares - discriminant analysis (PLS-DA), respectively.

3.4.1. Principal component analysis

Principal component analysis is a non-supervised multivariate technique that decomposes the data in a set of orthogonal variables called principal components (PCs). These PCs reduce the original data dimension and only contain the most valuable information. Each PC contains a determined variance of the original variables, the first PC (PC1) describes the maximum amount of the data variance, and the subsequent (PC2, PC3…) provide progressively less variance [31].

The representation of the sample coordinates of two PCs can be displayed in a two-dimension diagram named scores plot. Samples with similar properties shape clusters and the greater the difference, the greater the separation between them. Furthermore, a projection of the variables of the PCs can be employed to show how strongly each characteristic influences a PC. This representation is called loading plot and allows to obtain information of the most descriptive or discriminant variables.

A PCA model provides an exploratory analysis of all the information in a data table, offering an overview of the groupings, trends, patterns, and outliers of the studied samples [32]. For that reason, PCA is used in numerous applications as a first step [31].

3.4.2. Partial least squares - discriminant analysis

Partial least squares - discriminant analysis is a supervised classification method that relates the experimental data (here fingerprints, as the chromatograms of the samples) with their class membership. In PLS-DA converts the original data into linearly uncorrelated variables called latent variables (LVs) expressing the quantitative correlation between an X matrix (data fingerprints) and a Y matrix (class belonging).

This method seeks for the maximum covariance between these matrices and for that, a confusion matrix is generated with the proposed classification models. Then, a prediction of an "unknow set" is carried out in order to check the model performance, i.e. how many samples are correctly classified [31].

In order to obtain the best prediction in the classification - validation, the optimal number of LVs has to be chosen. Too few may retain insufficient information in the calibration, and too many may lead to overfitting so that the model is not robust in front of small variations. For that reason, the ideal number of LVs is the one with the minimum error in the model approach.

4. OBJECTIVES

This project aims to develop non-targeted methods based on high-performance liquid chromatography with fluorescence (HPLC-FLD) and ultraviolet (HPLC-UV) detection to achieve the characterization, classification, and authentication of paprika samples regarding their geographical indication. The next steps will be performed to achieve this objective:

1. Paprika samples will be submitted to a simple treatment to extract their bioactive compounds.
2. The gradient elution used in the chromatographic separation will be optimized based on a paprika sample analyzed by the HPLC-FLD method.
3. The obtained extracts will be analyzed employing HPLC-FLD and HPLC-UV to obtain the sample chromatographic fingerprints.
4. PCA will be applied to study the viability of HPLC-FLD and HPLC-UV fingerprints as sample chemical descriptors, the reproducibility of the method, and the robustness of the chemometric results.
5. PLS-DA will be employed to study the samples' classification according to their geographical origin and build a classification decision tree to calculate the sensitivity, specificity, and accuracy of both methods.

5. EXPERIMENTAL SECTION

5.1. REAGENTS AND SOLUTIONS

The solvents used for extraction in the sample treatment, as well as to optimize and carry out the chromatographic separation, were:

- Acetonitrile (ACN UHPLC Supergradient quality from Panreac)
- Formic acid (≥96% from Sigma-Aldrich)
- Methanol (99.9% from Panreac)
- Mili-Q water: Water was purified using an Elix 3 coupled to a Mili-Q system (Millipore, Bedford, MA, USA) and filtered through a 0.22 μm nylon filter integrated into the Mili-Q system.

5.2. INSTRUMENTATION AND CHROMATOGRAPHIC CONDITIONS

For the liquid chromatography separation, an Agilent 1100 Series HPLC instrument was employed. It was equipped with a binary pump (G1312A), a vacuum degasser (G1379A), an autosampler (G1367A), a diode array detector (G1315B) in series with a fluorescence detector (G1321A), and a computer with the Agilent ChemStation software.

The chromatographic separation was carried out using a Kinetex core-shell C18 column (100 mm x 4.6 mm i.d., 2.6 μm particle size) and guard column (2 mm × 4.6 mm i.d., 2.6 μm particle size), both from Phenomenex (Torrance, CA, USA), and a mobile phase composed of formic acid 0.1% (v/v) aqueous solution (solvent A) and acetonitrile (solvent B). Throughout the chromatographic run, the flow rate was established at 0.5 $mL \cdot min^{-1}$, and the used gradient elution is shown in Table 1. As seen, to obtain the best chromatographic separation, a combination of isocratic and lineal gradients at 3 different acetonitrile concentrations was the chosen option. Besides, an injection volume of 5 μL was set up for each sample.

Table 1. Gradient elution used in the chromatographic separation.

Time [min]	Solvent B [%]	Elution mode
0 - 2	40	Isocratic
2 - 3	40 - 80	Lineal
3 - 8	80	Isocratic
8 - 10	80 - 100	Lineal
10 - 12	100	Isocratic
12 - 13	100 - 40	Lineal
13 - 18	40	Isocratic

Regarding the fingerprint's detection, 280 nm was chosen as the absorbance wavelength for UV detection, while 310 and 380 nm were set as the excitation and emission wavelengths for FLD acquisition, respectively.

5.3. SAMPLES AND SAMPLE TREATMENT

A total of 122 paprika samples purchased from markets in Spain, Czech Republic and Hungary were analyzed. Figure 1 shows the number of samples from each region and the relative percentage of flavors within each one. For example, out of the 45 samples from La Vera, a third were hot flavors, so, 15 hot paprika samples from La Vera were studied.

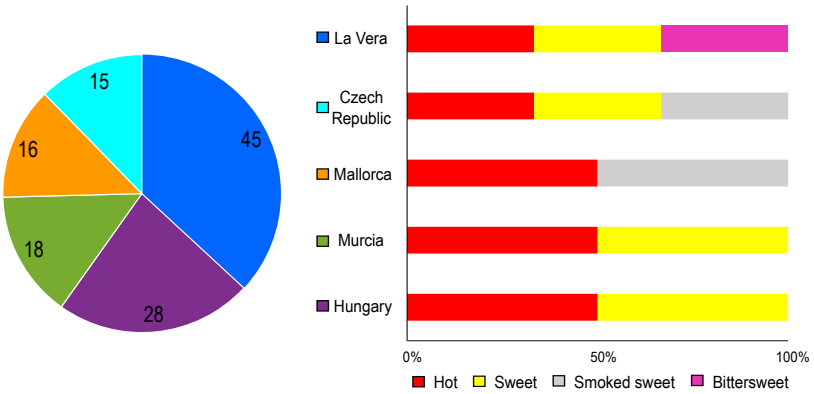

Figure 1. Pie chart of the number of samples of the 5 studied regions (La Vera, Czech Republic, Mallorca, Murcia and Hungary) and a 100% stacked bar chart with the respective flavors.

Sample treatment was carried out following a previously described method [18]. Briefly, 0.3 g of paprika were weighed in a 15 mL polypropylene tube, and 3 mL of water:acetonitrile (20:80 v/v) were added. Then, the mixture was stirred for 1 min in a Vortex (Stuart, Stone, United Kingdom), sonicated for 15 min (5510 Branson ultrasonic bath, Hampton, NH, USA), and centrifuged 30 min at 4500 rpm (ROTANTA 460 HR Centrifuge, Hettich, Germany). Finally, the resulting supernatant extract was filtered through a 0.22 µm nylon filter and preserved at 4ºC in a 2 mL glass injection vial until analysis.

To minimize the influence of instrumental drifts in the subsequent chemometric models, all samples were randomly analyzed. Moreover, a quality control (QC) sample, consisting of a mixture prepared with 50 µL of each sample extract, as well as an extracting solvent blank, were injected at the beginning and after every ten sample injections to control the repeatability of the method and cross-contamination during sample sequence, respectively.

5.4. DATA ANALYSIS

Once the samples were analyzed, the raw data obtained with the Agilent ChemStation software was exported to a spreadsheet using UniChrom™, from Seachrom Inc. (Lynnwood, WA, USA). At this point, all the chromatographic fingerprints could be represented and Solo, a stand-alone chemometrics software from Eigenvector Research (Manson, WA, USA), was used for the PCA and PLS-DA calculations.

In this study, both PCA and PLS-DA were used. While non-supervised PCA was used to observe the QC sample behavior and to study the samples' structure, supervised PLS-DA was employed for sample classificatory purposes. Indistinctly of the chemometric method used, the construction of different data matrices was required. The X-data matrix, which was used either in PCA and PLS-DA, consisted of the HPLC-UV (λ = 280 nm) or HPLC-FLD (λexc = 310 nm, λem = 380 nm) chromatographic fingerprints obtained. Besides, in PLS-DA, the class membership of each analyzed sample was defined in the Y-data matrix. In order to build the PLS-DA models, the first significant minimum point of the cross-validation (CV) error from the Venetian blind approach was chosen as the most appropriate number of LVs. Moreover, aiming at the improvement of data quality, the chromatographic fingerprints were smoothed, baseline-corrected, aligned, and autoscaled.

Considering the complexity of the studied issue, different PLS-DA models were consecutively combined constituting a classification decision tree, using the hierarchical model builder. The applicability of the chemometric method was then evaluated by external validation. First, PLS-DA calibration (CAL) models were created using 60% of the paprika samples (stratified random chosen). Instead, the remaining 40% was used as the validation (VAL) set.

6. RESULTS AND DISCUSSION

6.1. CHROMATOGRAPHIC SEPARATION

The objective of this work was to develop a HPLC-FLD method for the characterization, classification, and authentication of paprika according to their region of origin. In order to obtain the richest chromatographic fingerprints (i.e. with the maximum number of separated peaks), one sample from La Vera was employed for the optimization of the chromatographic separation in HPLC-FLD.

First, a mobile phase composed of formic acid 0.1% (v/v) aqueous solution and methanol was used in a simple gradient, as seen in Figure 2A. After 0.5 min at 5% organic phase, a linear gradient to 85% took place in 3.5 min and was followed by another linear gradient to 95% in 4 min and remained at that percentage one more min before returning to the initial conditions. The obtained chromatographic fingerprint is observed in Figure 2B.

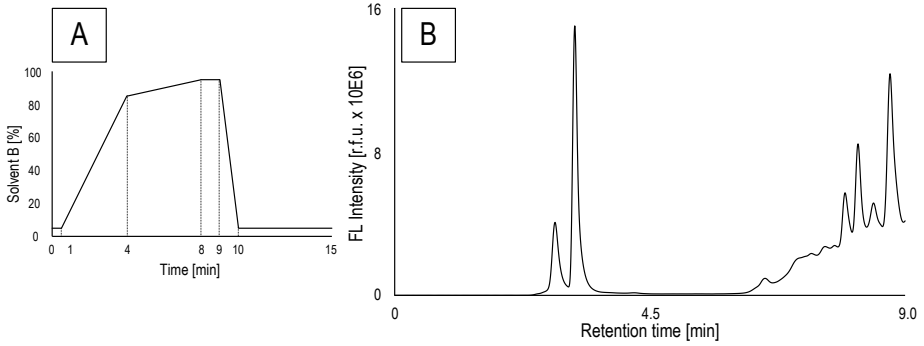

Figure 2. (A) First gradient elution tried in the chromatographic separation using methanol as solvent B and (B) the respective HPLC-FLD fingerprint (λexc = 310 nm, λem = 380 nm) of a sample from la Vera.

Since the compounds were grouped in the same area where the methanol was 95%, it was decided to use a less polar solvent with the purpose of having a better elution since the chromatographic separation was carried out with a reversed-phase column. Even employing

acetonitrile as solvent B with a simple linear gradient from 5% to 95%, no remarkable change in retention was observed and, for that reason, a different gradient elution program was proposed. In the beginning, a linear increase from 40% to 85% was done in 1.5 min. After 5.5 min at 85%, another increase to 100% was carried out in 2 min. The column was re-equilibrated after that percentage remained one more minute. A big improvement was then obtained in the chromatographic separation of paprika detected signals, especially in the retention time, but with the aim of obtaining more resolved and separated peaks, the gradient elution mentioned in Section 5.3 (Figure 3A) was chosen.

The chromatographic fingerprint obtained after extending the initial part with an isocratic 40% of solvent B (Figure 3B) had enough quality, at priori, for the objective of this work (untargeted method), so, no more improvements in the chromatographic fingerprint separation were tried. Since from 12 to 18 min the column was re-equilibrated at the initial conditions for the next chromatographic injection, in this work, fingerprints only encompass from 0 to 12 min.

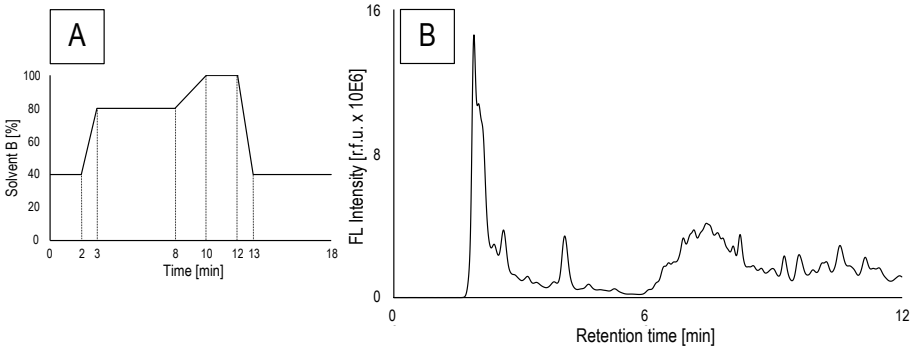

Figure 3. (A) Optimized gradient elution employed in the chromatographic separation using acetonitrile as Solvent B and (B) the respective HPLC-FLD fingerprint (λexc = 310 nm, λem = 380 nm) of a sample from la Vera.

6.2. CHROMATOGRAPHIC FINGERPRINTS

As previously mentioned, paprika samples from five different regions were analyzed by the proposed HPLC-FLD method for classification purposes. In the end, both HPLC-UV (λ = 280 nm) and HPLC-FLD (λexc = 310 nm, λem = 380 nm) chromatographic fingerprints were studied.

6.2.1. HPLC-FLD chromatographic fingerprints

To begin with, an HPLC-FLD fingerprint of a random spicy sample from each region is displayed in Figure 4.

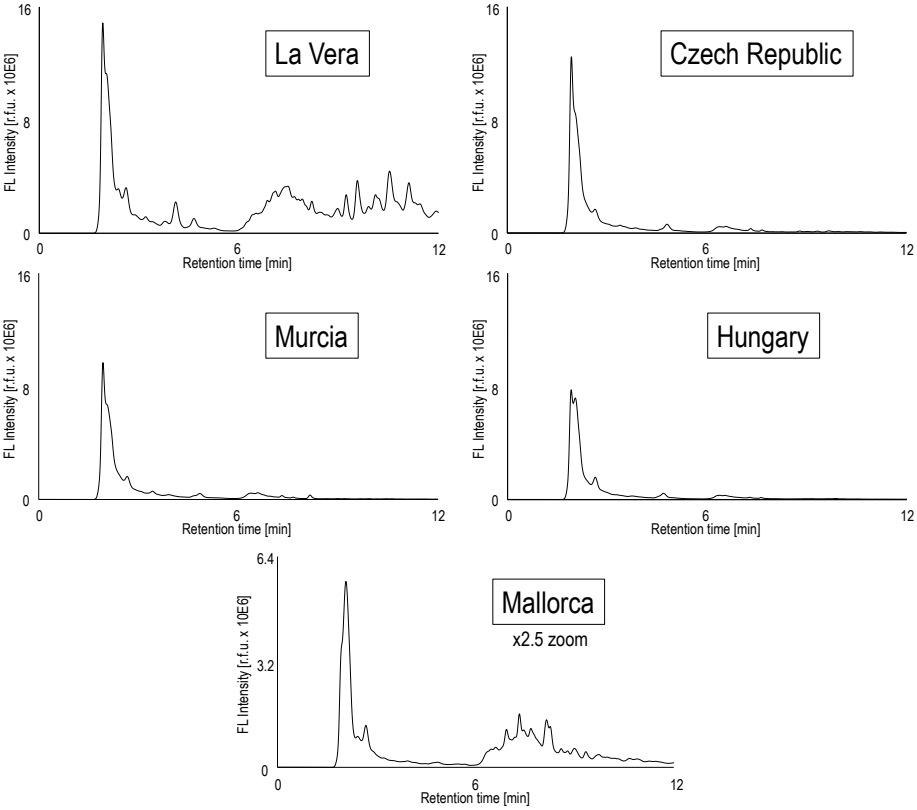

Figure 4. HPLC-FLD chromatographic fingerprints (λexc = 310 nm, λem = 380 nm) for a random spicy paprika sample within each region.

At a first glance, HPLC-FLD fingerprints from La Vera and Mallorca can be differentiated from those of the other regions without any problem. After the huge peak related to non-retained compounds, they are characterized by a richer area (regarding detected bioactive compounds and their relative abundances) from 6 to 12 min. HPLC-FLD fingerprints obtained for Czech Republic, Murcia, and Hungary were also similar, in fact, the detected compounds also appeared mainly in the same area and providing very similar peak signal distribution. Anyway, peak intensity variations among all the regions can be easily spotted, hence, the obtained

HPLC-FLD fingerprints were employed as sample chemical descriptors for the chemometric study.

6.2.2. HPLC-UV chromatographic fingerprints

The HPLC-UV fingerprints of the same samples above are shown in Figure 5.

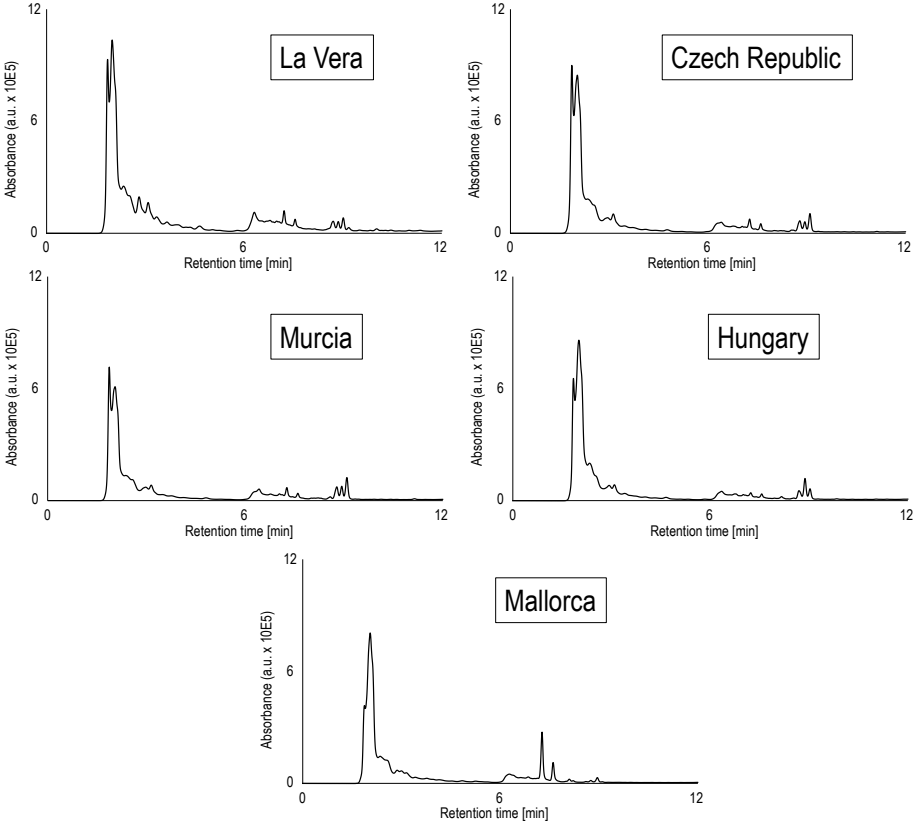

Figure 5. HPLC-UV chromatographic fingerprints (λ = 280 nm) for a selected spicy paprika sample within each region.

On this occasion, all the chromatographic fingerprints seem to have the same detected compounds, but with differences of relative peak intensities between regions. Again, as differences between the relative intensities among the analyzed paprika samples are observed, these HPLC-UV fingerprints were also evaluated as sample chemical descriptors for the chemometric classification of the analyzed samples.

6.3. EXPLORATORY STUDIES BY PCA

In order to evaluate the usefulness, repeatability, and robustness of HPLC-FLD method for paprika classification, a non-supervised exploratory PCA study was performed with the obtained HPLC-FLD and HPLC-UV fingerprints. As mentioned in Section 5.4, raw data was pretreated to reduce noise interferences, baseline drifts, and peak shifting. Furthermore, by decreasing differences in the magnitude scale, the same influence was granted to each variable [33].

6.3.1. HPLC-FLD study by PCA

With this objective in mind, a data matrix (135 × 1669) consisting of the recorded fluorescence intensity signals for each paprika sample and QCs was built, and subjected to PCA. The Samples vs. QCs plot of scores of PC1 vs. PC2 is shown in Figure 6.

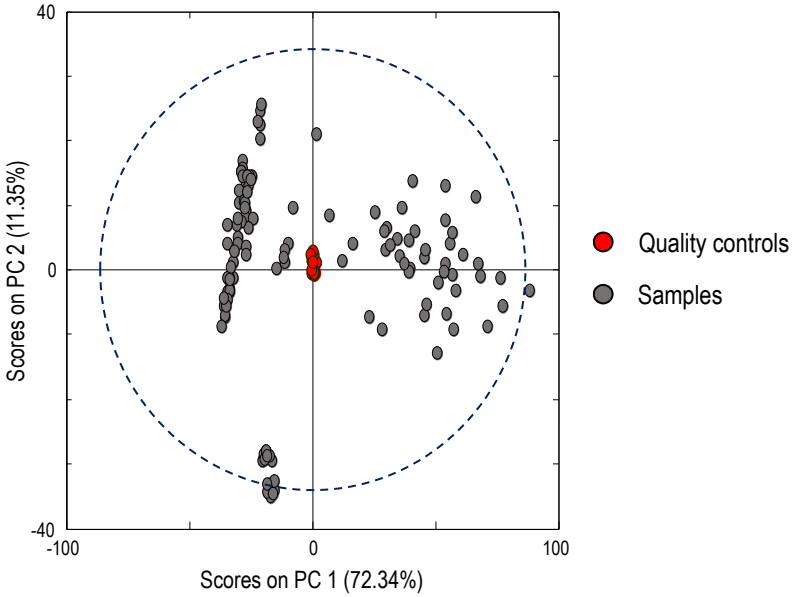

Figure 6. Scores plot of PC1 versus PC2 between QCs and paprika samples corresponding to HPLC-FLD fingerprints.

As it can be seen, QCs form a cluster, proving there were no experimental errors and the good performance of the proposed methodology, so, chemometric results were considered reliable. Thus, no correction was needed for the data matrix and an exploratory PCA distinguishing among regions was done. Figure 7 shows the obtained plot of scores PC1 vs. PC2.

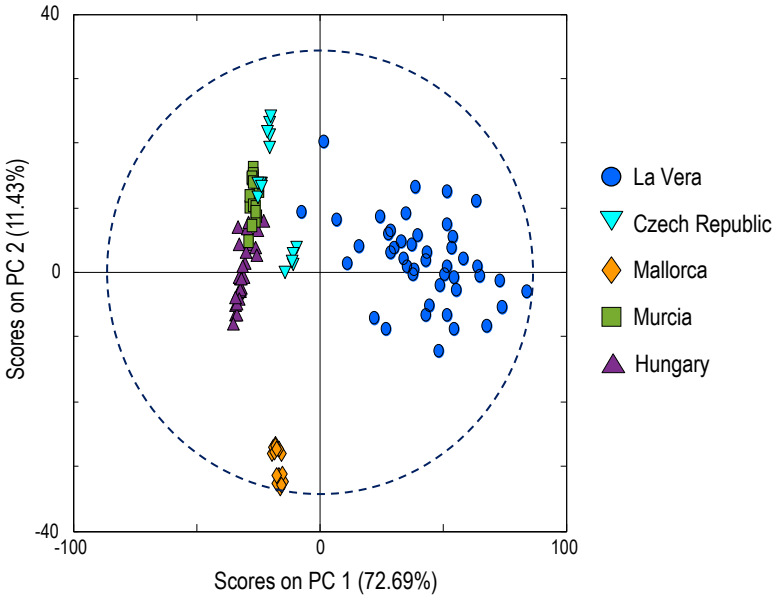

Figure 7. Scores plot of PC1 versus PC2 between regions and without QCs, corresponding to HPLC-FLD fingerprints.

As expected, La Vera and Mallorca can be easily distinguished from the others, probably due to the main differences observed in their HPLC-FLD fingerprints as commented in section 6.2.1. On the other hand, Czech Republic and Hungary paprika samples, overlap with Murcia samples, although samples tend to be grouped according to their production region. A peculiar thing is that among Czech Republic and Mallorca there are 3 and 2 differentiated groups, respectively, because this PCA has been able to differentiate between some sample flavors. Even without using a supervised method, it seems that the HPLC-FLD fingerprints are good enough to be used as sample chemical descriptors to address paprika classification and authentication.

6.3.2. HPLC-UV study by PCA

A data matrix (135 × 1669) with absorbance signals of HPLC-UV fingerprints was built, and the equivalent PCA scores plots as those previously described for HPLC-FLD are represented in Figure 8.

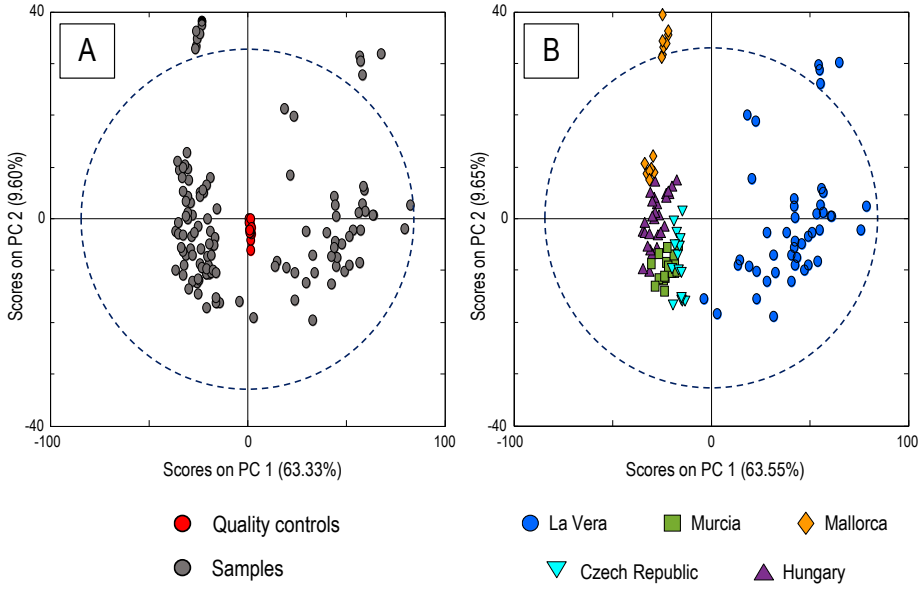

Figure 8. Scores plot of PC1 versus PC2 corresponding to HPLC-UV fingerprints (A) between QCs and paprika samples (B) between paprika regions and without QCs.

QCs also form a cluster, as expected, showing the good performance of the proposed methodologies. Regarding the grouping of samples according to their production region, HPLC-UV seems to provide worse results, especially for some paprika samples produced in Mallorca, which are now clustered close to Hungary paprika samples, which at the same time are clustered close to samples from Murcia and Czech Republic. For now, the proposed HPLC-FLD fingerprints seem to provide better results than the HPLC-UV ones. Nevertheless, as far as classification concerns, nothing can be concluded because, as told, PCA is only a non-supervised exploratory method.

6.4. CLASSIFICATION OF SAMPLES BY PLS-DA

With the aim of improving the PCA results and organize the classification decision tree, a supervised classification chemometric method such as PLS-DA was used.

6.4.1. HPLC-FLD classification by PLS-DA

As indicated in Section 5.4, a Y-data matrix (122 × 1670) was built removing the QCs and including the class (region) of each sample in the X-matrix employed in HPLC-FLD PCA. Then, a scores plot of LV1 vs. LV2 (five, the first minimum point of the CV error, was chosen as the number of LVs) was represented. It is observed in Figure 9.

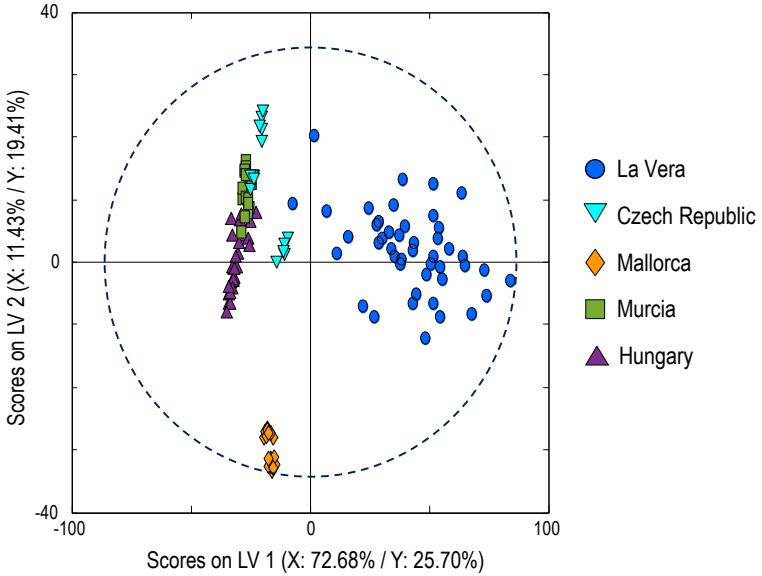

Figure 9. Scores plot of LV1 versus LV2 between regions, corresponding to HPLC-FLD fingerprints.

Unexpectedly, the PLS-DA scores plot shows no improvement regarding the exploratory PCA. Most probably, the obtained HPLC-FLD fingerprints are sufficiently discriminating so that an unsupervised method allows the correct discrimination without the need to specify the class.

To evaluate the classification capabilities of the proposed models, a classification plot of sample vs. Y-predicted class for each region against the others was represented using the CAL matrix (78 × 1670) explained in Section 5.4 as training set. The VAL matrix (48 × 1670) was employed for the prediction results, which were expressed as the proportion of the studied region samples that are correctly identified (sensitivity) and the proportion of correctly recognized as others (specificity). Table 2 shows the results of the predictions for the five regions, and Figure 10 serves as an example of the best and the worst obtained plots. In the

latter ones, the red mark represents the classification threshold line. Above this, samples are assigned to a predefined class and below to the opposite class.

Table 2. Sensitivity and specificity results for each region prediction in the first sequence.

Region	Sensitivity [%]	Specificity [%]
Czech Republic	100.0	95.2
Hungary	75.0	100.0
Mallorca	100.0	100.0
Murcia	100.0	81.0
La Vera	100.0	100.0

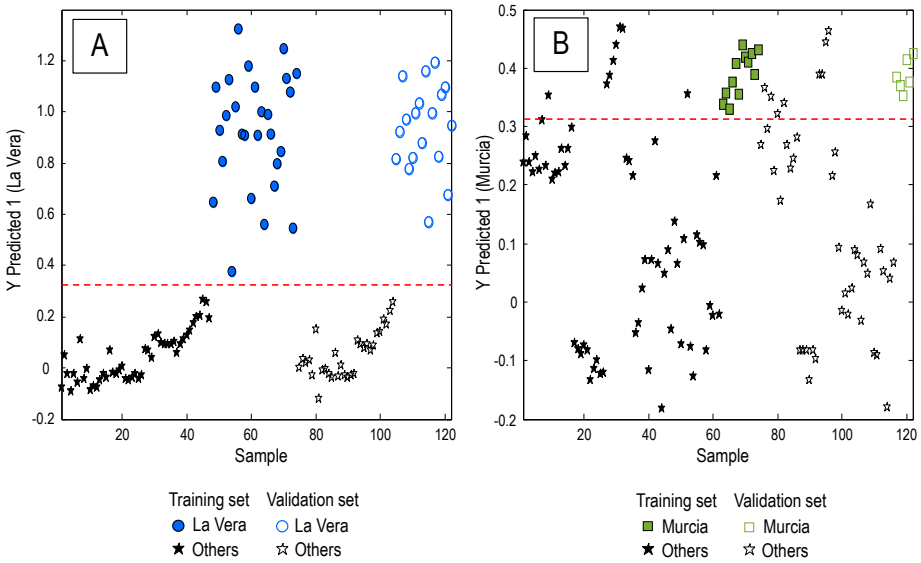

Figure 10. Sample versus Y predicted 1 scores plot with HPLC-FLD calibration and validation matrix for (A) La Vera vs. other regions and (B) Murcia vs. other regions.

Then, to build a classification decision tree, the region with the best prediction results was removed from the matrices. In case of a tie, the region with more replicates was considered better. In this case, La Vera prediction model was saved and the 45 samples were eliminated (27 from CAL matrix and 18 from VAL matrix), then, another set of predictions was done with the remaining regions. The process was repeated till only two regions remained. Table 3 and Table 4 show the obtained results for the last sets.

Table 3. Sensitivity and specificity results for each region prediction in the second sequence.

Region	Sensitivity [%]	Specificity [%]
Czech Republic	100.0	100
Hungary	91.6	100.0
Mallorca	100.0	100.0
Murcia	100.0	75.0

Table 4. Sensitivity and specificity results for each region prediction in the third sequence.

Region	Sensitivity [%]	Specificity [%]
Czech Republic	100.0	88.9
Hungary	83.3	100.0
Murcia	83.3	77.8

Mallorca and Czech Republic prediction models for the second and third sequence were saved respectively, as well as the model obtained from comparing Hungary and Murcia. Figure 11 illustrates the Y-predicted plots of the chosen models for the classification decision tree.

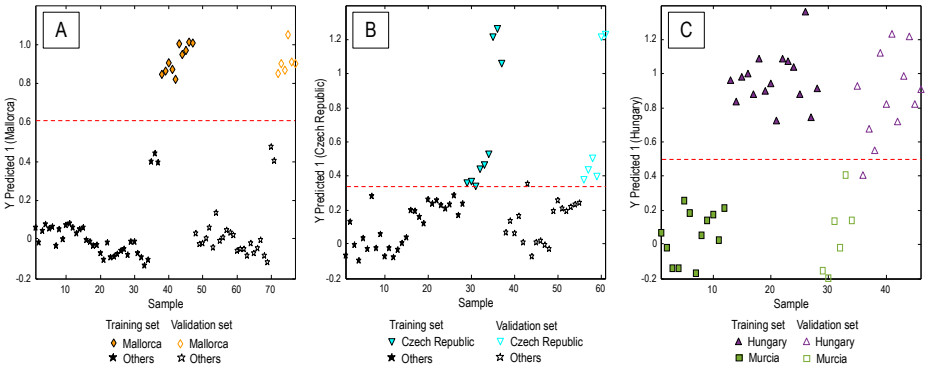

Figure 11. Sample versus Y predicted 1 scores plot with HPLC-FLD calibration and validation matrix for (A) Mallorca vs. other regions (without La Vera), (B) Czech Republic vs. other regions (without La vera and Mallorca), and (C) Hungary vs. Murcia.

6.4.2. HPLC-UV classification by PLS-DA

The classification of paprika samples by employing HPLC-UV fingerprints as chemical descriptors was also performed following the same process as previously described with HPLC-FLD fingerprints. In Figure 12 two PLS-DA scores plots (choosing six as the most appropriate number of LVs) are represented.

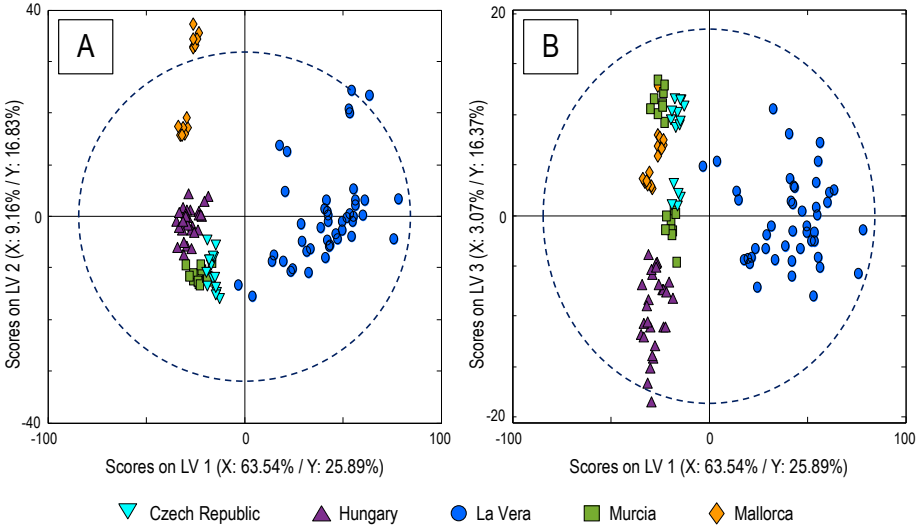

Figure 12. (A) Scores plot of LV1 versus LV2 and (B) LV1 versus LV3, both between regions and corresponding to HPLC-UV fingerprints.

In this case, and as expected, a great improvement over the PCA scores plot is appreciated. Also, since more than one comparison between LVs give a good region separation (Hungary is completely differentiated in LV1 vs. LV3 scores plot), it could be said that HPLC-UV fingerprints seem to provide better classifications than HPLC-FLD fingerprints. To make this sure, the classification capabilities using the HPLC-UV data was also carried out.

As seen in Figure 13, the Y-predicted plots seem similar to the above, for that reason, the same regions in the same order as HPLC-FLD were used as a model for the HPLC-UV to build a classification decision tree.

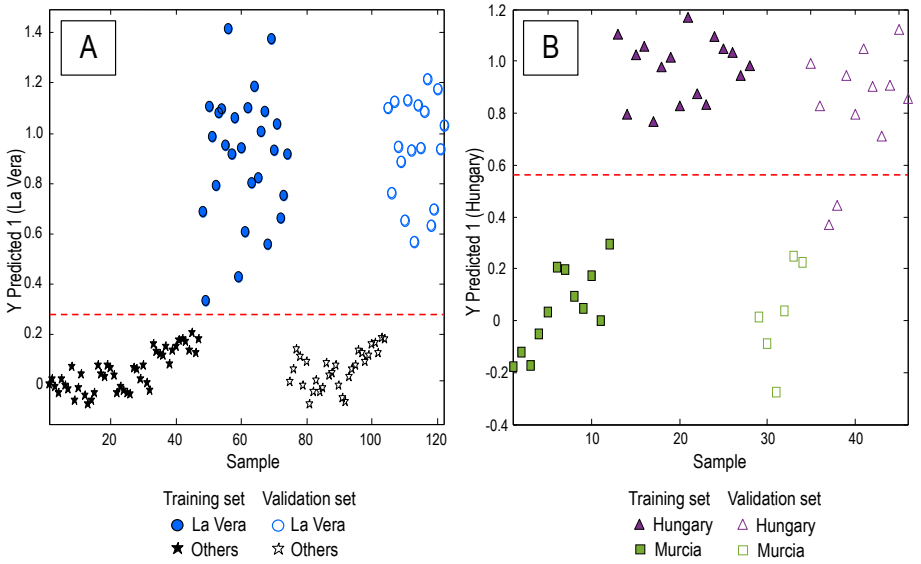

Figure 13. Sample versus Y predicted 1 scores plot with HPLC-UV calibration and validation matrix for (A) La Vera vs. other regions (first set) and (B) Hungary vs. Murcia (last set).

6.5. VALIDATION RESULTS

As previously commented, once the best models were chosen with the PLS-DA classification method, the decision tree was created with the help of the hierarchical model builder, using as rule the saved CAL models. Figure 14 illustrates the classification decision tree employed for the validation of the HPLC-FLD and HPLC-UV methods.

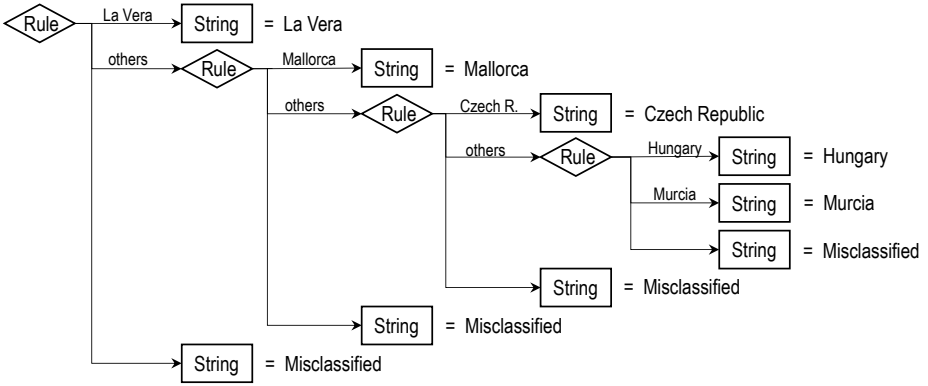

Figure 14. Classification decision tree employed for the categorization of the validation matrix of HPLC-FLD and HPLC-UV.

The obtained results after analyzing the HPLC-FLD VAL matrix with the built tree are observed in Table 5. As expected, sensitivity and specificity for La Vera and Mallorca were 100%. The classification PLS-DA was always perfect and, even in the exploratory HPLC-FLD PCA, these regions were completely differentiated from the others. Czech Republic, Murcia and Hungary had similar fingerprints and appeared grouped in the PCA and PLS-DA classification, so, the confusion between these samples was not strange. However, the prediction capabilities of the proposed HPLC-FLD fingerprinting method (specificity) is good, with values always higher than 97.6%.

Table 5: Prediction results for HPLC-FLD.

Region	Sensitivity [%]	Specificity [%]
Czech Republic	100.0	97.6
Hungary	83.3	100.0
Mallorca	100.0	100.0
Murcia	100.0	97.6
La Vera	100.0	100.0

Instead, the classification results with HPLC-UV fingerprints are shown in Table 6. Similar values as before were obtained, for the same reasons. Now, two samples from Hungary (not the same ones for both methods) were classified as Murcia. At first glance, it was strange because one of the PLS-DA scores plot distinguished Hungary from the other regions perfectly, but later, in the prediction validation (Figure 13B), it is clearly shown that these two samples were misinterpreted as Murcia ones. Again, the prediction capabilities of the proposed HPLC-UV fingerprinting method (specificity) is good, with values always higher than 95.2%.

Table 6: Prediction results for HPLC-UV.

Region	Sensitivity [%]	Specificity [%]
Czech Republic	100.0	100.0
Hungary	83.3	100.0
Mallorca	100.0	100.0
Murcia	100.0	95.2
La Vera	100.0	100.0

In both situations, only two samples from Hungary were incorrectly classified. Therefore, a good accuracy (95.8%) for both HPLC-FLD and HPLC-UV fingerprinting methods in the classification of paprika samples was achieved, demonstrating the feasibility of the proposed fingerprinting methodologies to address the characterization, classification, and authentication of paprika samples regarding their geographical indication.

7. CONCLUSIONS

In this work, non-targeted HPLC-FLD (λexc = 310 nm, λem = 380 nm) and HPLC-UV (λ = 280 nm) methods to achieve characterization, classification and authentication of paprika regarding their geographical indication were developed.

The optimized gradient elution employed in the chromatographic separation had enough quality to make the fingerprints obtained from both detection systems under study viable sample chemical descriptors to address sample classification by chemometrics. Also, the proposed methodologies were reproducible and robust according to the QCs behavior.

In reference to the results obtained in the classification of paprika samples, HPLC-FLD and HPLC-UV fingerprinting methods seem to have great accuracy, with 95.8% in both cases. However, the prediction capabilities of the HPLC-FLD fingerprints (specificity > 97.6%) are a little higher than the HPLC-UV ones (specificity > 95.2%). Even so, in the case of prioritizing one, HPLC-UV fingerprinting would be chosen since it is cheaper, and it can be found in many more laboratories than HPLC-FLD. Even in the scenario where the objective is to differentiate between paprika tastes from the same region, HPLC-UV detection would be selected. Although this data has not been discussed in the present work, preliminary results seem to indicate that the classification of paprika regarding their flavor by using HPLC-UV fingerprints has much better accuracy than with the HPLC-FLD ones.

Future work will focus on studying the adulteration of paprika with samples from another geographical origin by partial least squares regression.

8. REFERENCES

[1] W. Verbeke, "Agriculture and the food industry in the information age," *Eur. Rev. Agric. Econ.*, vol. 32, no. 3, pp. 347–368, 2005.

[2] G. P. Danezis, A. S. Tsagkaris, F. Camin, V. Brusic, and C. A. Georgiou, "Food authentication: Techniques, trends & emerging approaches," *TrAC - Trends Anal. Chem.*, vol. 85, pp. 123–132, 2016.

[3] J. C. Moore, J. Spink, and M. Lipp, "Development and Application of a Database of Food Ingredient Fraud and Economically Motivated Adulteration from 1980 to 2010," *Journal of Food Science*. 2012.

[4] N. Carranco, M. Farrés-Cebrián, J. Saurina, and O. Núñez, "Authentication and quantitation of fraud in extra virgin olive oils based on HPLC-UV fingerprinting and multivariate calibration," *Foods*, vol. 7, no. 4, 2018.

[5] C. V. Di Anibal, I. Ruisánchez, and M. P. Callao, "High-resolution 1H Nuclear Magnetic Resonance spectrometry combined with chemometric treatment to identify adulteration of culinary spices with Sudan dyes," *Food Chem.*, vol. 124, no. 3, pp. 1139–1145, 2011.

[6] Y. Liu, E. E. D. Todd, Q. Zhang, J. R. Shi, and X. J. Liu, "Recent developments in the detection of melamine," *Journal of Zhejiang University: Science B*. 2012.

[7] S. Medina, J. A. Pereira, P. Silva, R. Perestrelo, and J. S. Câmara, "Food fingerprints – A valuable tool to monitor food authenticity and safety," *Food Chem.*, vol. 278, no. July 2018, pp. 144–162, 2019.

[8] Consejo de la union europea, "Reglamento (CE) N 510/2006 del consejo de 20 de marzo de 2006 sobre la proteccion de las indicaciones geograficas y de las denominaciones de origen en los productos agricolas y alimenticios," *D. Of. la union Eur.*, no. 10, pp. 12–25, 2006.

[9] N. Serrano *et al.*, "Characterization and classification of Spanish paprika (Capsicum annuum L.) by liquid chromatography coupled to electrochemical detection with screen-printed carbon-based nanomaterials electrodes," *Talanta*, vol. 189, no. July, pp. 296–301, 2018.

[10] I. Bata-Vidács *et al.*, "Investigation of regional differences of the dominant microflora of spice paprika by molecular methods," *Food Control*, vol. 83, pp. 109–117, 2018.

[11] A. S. Antonio, L. S. M. Wiedemann, and V. F. Veiga Junior, "The genus: Capsicum: a phytochemical review of bioactive secondary metabolites," *RSC Adv.*, vol. 8, no. 45, pp. 25767–25784, 2018.

[12] S. Barbosa, G. Campmajó, J. Saurina, L. Puignou, and O. Núñez, “Determination of Phenolic Compounds in Paprika by Ultrahigh Performance Liquid Chromatography-Tandem Mass Spectrometry: Application to Product Designation of Origin Authentication by Chemometrics,” *J. Agric. Food Chem.*, vol. 68, no. 2, pp. 591–602, 2020.

[13] S. Barbosa, J. Saurina, L. Puignou, and O. Núñez, “Classification and authentication of paprika by UHPLC-HRMS fingerprinting and multivariate calibration methods (PCA and PLS-DA),” *Foods*, vol. 9, no. 4, 2020.

[14] S. Kim *et al.*, “Genome sequence of the hot pepper provides insights into the evolution of pungency in Capsicum species,” *Nat. Genet.*, vol. 46, no. 3, pp. 270–278, 2014.

[15] Y. Fiamegos, C. Dumitrascu, S. Papoci, and M. B. de la Calle, “Authentication of PDO paprika powder (Pimentón de la Vera) by multivariate analysis of the elemental fingerprint determined by ED-XRF. A feasibility study,” *Food Control*, vol. 120, p. 107496, 2021.

[16] M. Ghasemnezhad, M. Sherafati, and G. A. Payvast, “Variation in phenolic compounds, ascorbic acid and antioxidant activity of five coloured bell pepper (Capsicum annum) fruits at two different harvest times,” *J. Funct. Foods*, vol. 3, no. 1, pp. 44–49, 2011.

[17] J. Mateo, M. Aguirrezábal, C. Domínguez, and J. M. Zumalacárregui, “Volatile Compounds in Spanish Paprika,” *J. Food Compos. Anal.*, vol. 10, no. 3, pp. 225–232, 1997.

[18] X. Cetó *et al.*, “Determination of HPLC-UV fingerprints of spanish paprika (Capsicum annuum L.) for its classification by linear discriminant analysis,” *Sensors (Switzerland)*, vol. 18, no. 12, 2018.

[19] O. Fayos *et al.*, “Assessment of Capsaicinoid and Capsinoid Accumulation Patterns during Fruit Development in Three Chili Pepper Genotypes (Capsicum spp.) Carrying Pun1 and pAMT Alleles Related to Pungency,” *J. Agric. Food Chem.*, vol. 67, no. 44, pp. 12219–12227, 2019.

[20] L. M. Reid, C. P. O’Donnell, and G. Downey, “Recent technological advances for the determination of food authenticity,” *Trends Food Sci. Technol.*, vol. 17, no. 7, pp. 344–353, 2006.

[21] O. Núñez and P. Lucci, “Application of liquid chromatography in food analysis,” *Foods*, vol. 9, no. 9, pp. 1–5, 2020.

[22] A. Palacios-Morillo, J. M. Jurado, A. Alcázar, and F. Pablos, “Differentiation of Spanish paprika from Protected Designation of Origin based on color measurements and pattern recognition,” *Food Control*, vol. 62, pp. 243–249, 2016.

[23] O. Monago-Maraña, C. E. Eskildsen, T. Galeano-Díaz, A. Muñoz de la Peña, and J. P. Wold, “Untargeted classification for paprika powder authentication using visible – Near infrared spectroscopy (VIS-NIRS),” *Food Control*, vol. 121, no. June 2020, 2021.

[24] B. Horn, S. Esslinger, M. Pfister, C. Fauhl-Hassek, and J. Riedl, “Non-targeted detection of paprika adulteration using mid-infrared spectroscopy and one-class

classification – Is it data preprocessing that makes the performance?," *Food Chem.*, vol. 257, no. March, pp. 112–119, 2018.

[25] S. Lohumi, H. Lee, M. S. Kim, J. Qin, and B. K. Cho, "Raman imaging for the detection of adulterants in paprika powder: A comparison of data analysis methods," *Appl. Sci.*, vol. 8, no. 4, 2018.

[26] Y. Hu, S. Wang, S. Wang, and X. Lu, "Application of nuclear magnetic resonance spectroscopy in food adulteration determination: The example of Sudan dye i in paprika powder," *Sci. Rep.*, vol. 7, no. 1, pp. 1–9, 2017.

[27] X. Cetó, C. Sánchez, N. Serrano, J. M. Díaz-Cruz, and O. Núñez, "Authentication of paprika using HPLC-UV fingerprints," *Lwt*, vol. 124, no. August 2019, 2020.

[28] N. Z. Ballin and K. H. Laursen, "To target or not to target? Definitions and nomenclature for targeted versus non-targeted analytical food authentication," *Trends Food Sci. Technol.*, vol. 86, no. September 2018, pp. 537–543, 2019.

[29] L. Cuadros-Rodríguez, C. Ruiz-Samblás, L. Valverde-Som, E. Pérez-Castaño, and A. González-Casado, "Chromatographic fingerprinting: An innovative approach for food 'identitation' and food authentication - A tutorial," *Anal. Chim. Acta*, vol. 909, pp. 9–23, 2016.

[30] S. Esslinger, J. Riedl, and C. Fauhl-Hassek, "Potential and limitations of non-targeted fingerprinting for authentication of food in official control," *Food Res. Int.*, vol. 60, pp. 189–204, 2014.

[31] D. Granato *et al.*, "Trends in Chemometrics: Food Authentication, Microbiology, and Effects of Processing," *Compr. Rev. Food Sci. Food Saf.*, vol. 17, no. 3, pp. 663–677, 2018.

[32] J. Trygg, E. Holmes, and T. Lundstedt, "Chemometrics in metabonomics," *Journal of Proteome Research*. 2007.

[33] G. Campmajó, L. Cayero, J. Saurina, and O. Núñez, "Classification of Hen Eggs by HPLC-UV fingerprinting and chemometric methods," *Foods*, vol. 8, no. 8, pp. 1–10, 2019.

9. ACRONYMS

ACN: Acetonitrile

a.u.: Arbitrary Units

CAL: Calibration

CV: Cross-Validation

EC: Council Regulation

EFSA: European Food Safety Authority

FL: Fluorescence

HPLC: High-Performance Liquid Chromatography

HPLC-FLD: High-Performance Liquid Chromatography with fluorescence detection

HPLC-UV: High-Performance Liquid Chromatography with ultraviolet detection

i.d.: Internal Diameter

ISO: International Standardization Organization

LC-HRMS: Liquid Chromatography coupled to High-Resolution Mass Spectrometry

LC-MS: Liquid Chromatography coupled to Mass Spectrometry

LC-UV: Liquid Chromatography with ultraviolet detection

LV: Latent Variable

NMR: Nuclear Magnetic Resonance

PC: Principal Component

PCA: Principal Component Analysis

PDO: Protected Designation of Origin

PLS-DA: Partial Least Squares - Discriminant Analysis

QC: Quality Control

r.f.u.: Relative Fluorescence Units

UHPLC: Ultra High-Performance Liquid Chromatography

UV-VIS: Ultraviolet-Visible

VAL: Validation